

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-45289

(43) 公開日 平成11年(1999) 2月16日

(51) Int.Cl.⁴
G 0 6 F 17/30
17/27

識別記号

F I
G 0 6 F 15/401 3 2 0 A
15/20 5 5 0 A
15/40 3 7 0 A
15/403 3 4 0 A

審査請求 未請求 請求項の数 7 F D (全 12 頁)

(21) 出願番号 特願平9-218230

(22) 出願日 平成9年(1997) 7月28日

(71) 出願人 390024350

株式会社ジャストシステム

徳島県徳島市神浜東3-46

(72) 発明者 野村 直之

徳島県徳島市神浜東3丁目46番地 株式会
社ジャストシステム内

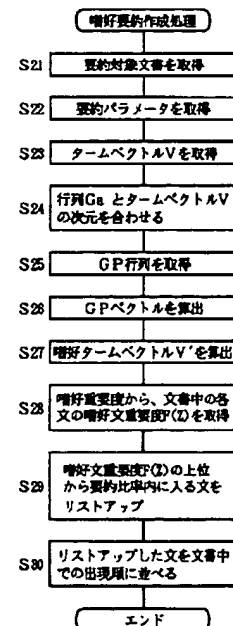
(74) 代理人 弁理士 川井 隆 (外1名)

(54) 【発明の名称】 文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法

(57) 【要約】

【課題】 本発明は、利用目的等のユーザーの嗜好を踏まえた要約の作成が可能な、文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法を提供すること。

【解決手段】 過去の処理文書の内容からキーワードとその重要度を取得し、複数のユーザーとキーワードとの一方を行、他方を列として前記各ユーザーに対する各キーワードの重要度を要素値とするGP行列を取得する。要約作成対象文書から重要語 a, b, ...と、その出現頻度等からの重要度を取得し、この重要度を要素としたタームベクトルVを、GP行列によってシフトさせ、嗜好タームベクトルV'を取得する。嗜好タームベクトルV'の要素(=嗜好重要度)をもとに要約作成対象文書から嗜好重要文F(Z)を抽出し、要約作成対象文書における出現順に並べて、嗜好要約とする。



【特許請求の範囲】

【請求項1】 複数の文よりなる文書を取得する文書取得手段と、

前記文書取得手段により取得された前記文書から重要語句とその重要度を取得する重要語句抽出手段と、

前記重要語句に基づいて前記文書からユーザーの嗜好を反映した嗜好重要部分を選択する嗜好重要部分選択手段と、

前記嗜好重要部分選択手段により選択された嗜好重要部分に基づいて前記文書の要約を作成する嗜好要約作成手段とを具備することを特徴とする文書処理装置。

【請求項2】 前記重要語句抽出手段は、

前記文書取得手段により取得された前記文書から前記重要語句の候補語句とその重要度を取得する候補語句取得手段と、

ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと各ユーザーの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得手段と、を有し、

前記嗜好取得手段により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記候補語句取得手段により取得された候補語句の重要度をシフトさせた重要度から前記重要語句を抽出し、

前記嗜好重要部分選択手段は、前記重要語句とその重要度により前記嗜好重要部分を選択することを特徴とする請求項1に記載の文書処理装置。

【請求項3】 前記重要語句抽出手段は、前記文書取得手段により取得された前記文書から前記重要語句の候補語句とその重要度を取得して、前記候補語句の重要度により前記重要語句を抽出し、

前記嗜好重要部分選択手段は、ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと複数のユーザーそれぞれの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得手段を有し、前記嗜好取得手段により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記重要語句抽出手段により取得された重要語句の重要度をシフトさせた重要度により前記重要部分を選択することを特徴とする請求項1に記載の文書処理装置。

【請求項4】 複数の文よりなる文書を取得する文書取得機能と、

前記文書取得機能により取得された前記文書から重要語句とその重要度を取得する重要語句抽出機能と、

前記重要語句に基づいて前記文書からユーザーの嗜好を反映した嗜好重要部分を選択する嗜好重要部分選択機能と、

前記嗜好重要部分選択機能により選択された嗜好重要部分に基づいて前記文書の要約を作成する嗜好要約作成機能とをコンピュータに実現させるためのコンピュータ読みとり可能な文書処理プログラムが記憶された記憶媒体。

【請求項5】 前記重要語句抽出機能は、

前記文書取得機能により取得された前記文書から前記重要語句の候補語句とその重要度を取得する候補語句取得機能と、

10 ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと各ユーザーの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得機能と、を有し、

前記嗜好取得機能により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記候補語句取得機能により取得された候補語句の重要度をシフトさせた重要度から前記重要語句を抽出し、

20 前記嗜好重要部分選択機能は、前記重要語句とその重要度により前記嗜好重要部分を選択することを特徴とする請求項4に記載した文書処理プログラムが記憶された記憶媒体。

【請求項6】 前記重要語句抽出機能は、前記文書取得機能により取得された前記文書から前記重要語句の候補語句とその重要度を取得して、前記候補語句の重要度により前記重要語句を抽出し、

前記嗜好重要部分選択機能は、ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと複数のユーザーそれぞれの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得機能を有し、前記嗜好取得機能により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記重要語句抽出機能により取得された重要語句の重要度をシフトさせた重要度により前記重要部分を選択することを特徴とする請求項4に記載した文書処理プログラムが記憶された記憶媒体。

40 【請求項7】 複数の文よりなる文書を取得し、取得された前記文書から重要語句とその重要度を取得し、

前記重要語句に基づいて前記文書からユーザーの嗜好を反映した嗜好重要部分を選択し、

選択された前記嗜好重要部分に基づいて前記文書の要約を作成することを特徴とする文書処理方法。

【発明の詳細な説明】

【0001】

50 【発明の属する技術分野】本発明は、文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理

方法に関し、更に詳細には、利用目的等のユーザーの嗜好を踏まえた要約の作成に関する。

【0002】

【従来の技術】従来、書籍、論文、報告書等の各種の文書に対し、要約（抄録を含む）の自動作成処理をコンピュータを用いて行うことが行われている。文書の自動要約については、例えば、「全文情報からの意味的情報の抽出と加工」（情報処理学会第38回全国大会予稿集、第222頁；1989年）で提案されている。この方法では、まず文書中の重要語を字種や動詞等の情報から抽出し、さらに重要語の出現頻度から最重要語を取得する。次に重要語と最重要語が出現するか否かから重要文を取得することで、自動的に要約を作成することが可能になる。また、文章の段落の性質を反映させることで、より正確に要約を作成する特開平3-191475号公報に記載された方法等も提案されている。

【0003】

【発明が解決しようとする課題】しかし、同一の文書でも、例えば営業用や技術資料用等の利用目的その他のユーザーの嗜好が異なると、文書における重要部位等に差異が生じる。そして、上述のような従来の文書処理によって要約を作成しても、ユーザーの嗜好を踏まえた要約を得ることはできない問題点がある。

【0004】本発明は、上述のような課題を解決するためになされたもので、利用目的等のユーザーの嗜好を踏まえた要約自動作成処理文書処理を行うことのできる文書処理装置、文書処理プログラムを記憶した記憶媒体、及び文書処理方法を提供することを目的とする。

【0005】

【課題を解決するための手段】請求項1に記載の発明は、複数の文よりなる文書を取得する文書取得手段と、前記文書取得手段により取得された前記文書から重要語句とその重要度を取得する重要語句抽出手段と、前記重要語句に基づいて前記文書からユーザーの嗜好を反映した嗜好重要部分を選択する嗜好重要部分選択手段と、前記嗜好重要部分選択手段により選択された嗜好重要部分に基づいて前記文書の要約を作成する嗜好要約作成手段と、を具備する文書処理装置を提供することにより、上記目的を達成する。請求項2に記載の発明は、請求項1に記載の文書処理装置において、前記重要語句抽出手段は、前記文書取得手段により取得された前記文書から前記重要語の候補語句とその重要度を取得する候補語句取得手段と、ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと各ユーザーの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得手段と、を有し、前記嗜好取得手段により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記候補語句取得手段により取得された候補語句の重要度をシフトさせた重要度から前記重要語句を抽出し、前記嗜好重要部分選択手段は、前記重要語句とその重要度により前記嗜好重要部分を選択する文書処理プログラムが記憶された記憶媒体を提供することにより前記目的を達成する。請求項6に記載の発明は、請求項4に記載の記憶媒体において、前記重要語句抽出機能は、前記文書取得機能により取得された前記文書から前記重要語の候補語句とその重要度を取得する候補語句取得機能と、ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと各ユーザーの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得機能と、を有し、前記嗜好取得機能により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記候補語句取得機能により取得された候補語句の重要度をシフトさせた重要度から前記重要語句を抽出し、前記嗜好重要部分選択機能は、前記重要語句とその重要度により前記嗜好重要部分を選択する文書処理プログラムが記憶された記憶媒体を提供することにより前記目的を達成する。請求項6に記載の発明は、請求項4に記載の記憶媒体において、前記重要語句抽出機能は、前記文書取得機能により取得された前記文書から前記重要語の候補語句とその重要度を取得して、前記候補語句の重要度により前記重要語句を抽出し、前記嗜好重要部分選択機能は、ユーザー

重要度をシフトさせた重要度から前記重要語句を抽出し、前記嗜好重要部分選択手段は、前記重要語句とその重要度により前記嗜好重要部分を選択することを文書処理装置を提供することにより、上記目的を達成する。請求項3に記載の発明は、請求項1に記載の発明において、前記重要語句抽出手段は、前記文書取得手段により取得された前記文書から前記重要語の候補語句とその重要度を取得して、前記候補語句の重要度により前記重要語句を抽出し、前記嗜好重要部分選択手段は、ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと複数のユーザーそれぞれの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得手段を有し、前記嗜好取得手段により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記重要語句抽出手段により取得された重要語句の重要度をシフトさせた重要度により前記重要部分を選択する文書処理装置を提供することにより前記目的を達成する。請求項4に記載の発明は、複数の文よりなる文書を取得する文書取得機能と、前記文書取得機能により取得された前記文書から重要語句とその重要度を取得する重要語句抽出機能と、前記重要語句に基づいて前記文書からユーザーの嗜好を反映した嗜好重要部分を選択する嗜好重要部分選択機能と、前記嗜好重要部分選択機能により選択された嗜好重要部分に基づいて前記文書の要約を作成する嗜好要約作成機能とをコンピュータに実現させるためのコンピュータ読みとり可能な文書処理プログラムが記憶された記憶媒体を提供することにより上記目的を達成する。請求項5に記載の発明は、請求項4に記載の記憶媒体において、前記重要語句抽出機能は、前記文書取得機能により取得された前記文書から前記重要語の候補語句とその重要度を取得する候補語句取得機能と、ユーザーの嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと各ユーザーの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得機能と、を有し、前記嗜好取得機能により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記候補語句取得機能により取得された候補語句の重要度をシフトさせた重要度から前記重要語句を抽出し、前記嗜好重要部分選択機能は、前記重要語句とその重要度により前記嗜好重要部分を選択する文書処理プログラムが記憶された記憶媒体を提供することにより前記目的を達成する。請求項6に記載の発明は、請求項4に記載の記憶媒体において、前記重要語句抽出機能は、前記文書取得機能により取得された前記文書から前記重要語の候補語句とその重要度を取得して、前記候補語句の重要度により前記重要語句を抽出し、前記嗜好重要部分選択機能は、ユーザー

の嗜好を表す複数のキーワードの重要度を要素値とする嗜好ベクトル、または、複数のユーザーと複数のユーザーそれぞれの嗜好を表す複数のキーワードとの一方を行、他方を列として前記各ユーザーに対する前記各キーワードの重要度を要素値とするGP行列、を取得する嗜好取得機能を有し、前記嗜好取得機能により取得された前記嗜好ベクトルまたは前記GP行列を用いて、前記重要語句抽出機能により取得された重要語句の重要度をシフトさせた重要度により前記重要部分を選択する文書処理プログラムが記憶された記憶媒体を提供することにより前記目的を達成する。請求項7に記載の発明は、複数の文よりなる文書を取得し、取得された前記文書から重要語句とその重要度を取得し、前記重要語句に基づいて前記文書からユーザーの嗜好を反映した嗜好重要部分を選択し、選択された前記嗜好重要部分に基づいて前記文書の要約を作成する文書処理方法を提供することにより前記目的を達成する。

【0006】

【発明の実施の形態】以下、本発明の文書処理装置、文書処理プログラムが記憶された記憶媒体、及び文書処理方法の好適な実施の形態について、図1から図7を参照して詳細に説明する。

(1) 実施形態の概要

本実施形態では、過去の処理文書の内容からキーワードとその重要度を取得し、複数のユーザーとキーワードとの一方を行、他方を列として前記各ユーザーに対する各キーワードの重要度を要素値とするGP行列を取得する。要約作成対象文書から重要語a、b、…と、その出現頻度等からの重要度 $g(p)$ 、 $g(q)$ 、…を取得し、重要度を要素としたタームベクトル $V = (g(p), g(q), \dots)$ を、GP行列によってシフトさせ、嗜好タームベクトル V' を取得する。嗜好タームベクトル V' の要素(=嗜好重要度) $g'(p)$ 、 $g'(q)$ 、…をもとに要約作成対象文書から嗜好重要文を抽出し、要約作成対象文書における出現順に並べて、嗜好要約とする。

【0007】(2)-実施形態の詳細

図1は、本発明の文書処理装置の一実施形態であり、本発明の文書処理プログラムを記憶した記憶媒体の一実施形態の該プログラムが読み取られたコンピュータの構成を表したブロック図である。この図1に示すように、文書処理装置(コンピュータ)は、装置全体を制御するための制御部11を備えている。この制御部11には、データバス等のバスライン21を介して、入力装置としてのキーボード12やマウス13、表示装置14、印刷装置15、記憶装置16、記憶媒体駆動装置17、通信制御装置18、および、入出力I/F19、および、文字認識装置20が接続されている。制御部11は、CPU111、ROM112、RAM113を備えている。ROM112は、CPU111が各種制御や演算を行うた

めの各種プログラムやデータが予め格納されたリードオンリーメモリである。

【0008】RAM113は、CPU111にワーキングメモリとして使用されるランダムアクセスメモリである。このRAM113には、本実施形態による嗜好要約処理を行うためのエリアとして、対象文書格納エリア1131、要約パラメータ格納エリア1132、重要語格納エリア1133、タームベクトル格納エリア1134、行列格納エリア1135、嗜好タームベクトル格納エリア1136、要約格納エリア1137、その他の各種エリアが確保されるようになっている。

【0009】対象文書格納エリア1131には、嗜好要約作成の対象となる文書が格納される。要約パラメータ格納エリア1132には、操作者からの入力等により取得された要約パラメータの値または後述のデータ格納部163から読み込んだ要約パラメータのデフォルト値が格納される。操作者が入力する要約パラメータとしては、例えば、全文書に対する要約の比率(1~99%)や、日付時刻、価格情報、物理量(サイズ、重量、温度等)等の数量優先のある/なし、URL(Uniform Resource Locator)重視長単文の優先のある/なし、です/ます/であるの選択をする/しない、等の値が格納される。タームベクトル格納エリア1134には、本実施形態により取得された、嗜好要約作成の対象文書の、タームベクトルが格納される。要約格納エリア1135には、本実施形態により取得された重要文が、嗜好要約作成対象文書における順番で格納される。

【0010】キーボード12は、かな文字を入力するためのかなキーやテンキー、各種機能を実行するための機能キー、カーソルキー、等の各種キーが配置されている。マウス13は、ポインティングデバイスであり、表示装置14に表示されたキーやアイコン等を左クリックすることで対応する機能の指定を行う入力装置である。表示装置14は、例えばCRTや液晶ディスプレイ等が使用される。この表示装置14には、嗜好要約作成の対象となる文書の内容や、本実施形態により作成された嗜好要約等が表示されるようになっている。印刷装置15は、表示装置14に表示された文章や、記憶装置16の文書データベース165に格納された文書等の印刷を行うためのものである。この印刷装置としては、レーザプリンタ、ドットプリンタ、インクジェットプリンタ、ページプリンタ、感熱式プリンタ、熱転写式プリンタ、等の各種印刷装置が使用される。

【0011】記憶装置16は、読み書き可能な記憶媒体と、その記憶媒体に対してプログラムやデータ等の各種情報を読み書きするための駆動装置で構成されている。この記憶装置16に使用される記憶媒体としては、主としてハードディスクが使用されるが、後述の記憶媒体駆動装置17で使用される各種記憶媒体のうちの読み書き可能な記憶媒体を使用するようにしてもよい。記憶装置